

# Kennis in veelvoud

Rede

uitgesproken bij de aanvaarding van het ambt  
van hoogleraar in de Documentaire Informatiewetenschap  
aan de Universiteit van Amsterdam  
op woensdag 24 juni 1998

door

John Mackenzie Owen

Mijnheer de Rector Magnificus,  
Overige leden van het College van Bestuur,  
Waarde collega's,  
Dames en heren studenten,  
Zeer gewaardeerde toehoorders.

### **De toekomst van het heden is onzeker**

In het jaar 1986 publiceerde ik een boek over de toepassing van computers in bibliotheken<sup>1</sup>. Waarschijnlijk is dit een bijzonder boek. Niet zozeer vanwege de inhoud, maar omdat het een van de eerste Nederlandse boeken moet zijn waarbij het gehele proces voorafgaand aan het drukken digitaal, zonder gebruik van papier, is uitgevoerd. Het manuscript van dat boek bevindt zich op een magneetband; aantekeningen en eerste schetsen bevinden zich op een aantal diskettes. Dit materiaal heb ik altijd trouw bewaard, maar er is wel een probleem mee. Noch ik, noch iemand anders zal ooit in staat zijn dit materiaal te raadplegen. De apparatuur die nodig is voor het lezen van de fysieke informatiedragers is niet meer te koop. De tekstverwerkers en besturingssystemen waarmee het materiaal is geproduceerd, zijn niet meer beschikbaar. En als er ergens nog een computermuseum is met werkende apparatuur, dan zal blijken dat de magnetische patronen waarmee de informatie was gecodeerd, inmiddels van magneetband en diskettes zijn verdwenen. Al na ruim tien jaar is de oorspronkelijke - digitale - bron van het boek verdwenen naar het land van ooit en nergens. Gelukkig is het boek destijds nog gedrukt, en bevinden zich verschillende exemplaren in bibliotheken en bij mij thuis inde boekenkast.

In de gedrukte tekst van deze lezing vindt u een aantal verwijzingen naar digitale bronnen op het Internet. U kunt met behulp van een 'browser' zelf deze bronnen raadplegen. Nog wel. Want als u over twee jaar deze tekst weer eens ter hand neemt, is het zeer onwaarschijnlijk dat de aangegeven bronnen nog te vinden zijn. Misschien zijn ze tegen die tijd van plaats veranderd, waardoor u ze niet meer kunt lokaliseren. Waarschijnlijker is het dat de inhoud intussen aanzienlijk is aangepast. Maar het meest waarschijnlijk is het dat de door mij aangegeven bronnen letterlijk zijn verdwenen: geschrapt van de enige plaats waar ze te vinden waren, nergens gearchiveerd, voor altijd verdwenen.

De productie en distributie van publicaties, of liever het proces van vastleggen en verspreiden van kennis, is door digitalisering en het gebruik van netwerken enorm aan het veranderen. De voordelen van de informatietechnologie op dit terrein zijn bekend: snellere distributie, betere toegankelijkheid, potentieel lagere kosten. Met de twee voorbeelden die ik zo-even gaf, wil ik u echter duidelijk maken dat er aan het gebruik van informatietechnologie voor de verspreiding van kennis ook nadelen verbonden zijn.

Heel kort samengevat, kunnen we het probleem als volgt beschrijven. Digitale informatie bestaat doorgaans in maar één exemplaar, op één locatie op het netwerk. Meer exemplaren zijn niet nodig: het is immers het wezen van het netwerk dat iedereen altijd, waar hij of zij zich ook bevindt, bij dat ene exemplaar kan komen. Maar dat ene exemplaar bestaat slechts zolang de eigenaar ervan bereid is het er te laten staan, er niets aan te veranderen, en het via het netwerk toegankelijk te houden. De eigenaren en beheerders van informatie op het netwerk, of dat nu auteurs, uitgevers, bibliotheken of anderen zijn, blijken daar slechts voor een zeer korte periode toe bereid. Het eigen belang bij langdurige archivering is gering - of wordt als gering ervaren.

Digitale informatie op het netwerk moet daarom door anderen worden gearchiveerd.<sup>2</sup> Het is in de praktijk al niet zo eenvoudig om dat te organiseren, en het

gevolg is dan ook dat zeer veel informatie momenteel verloren gaat. Maar mijn eerste voorbeeld gaf al aan dat archivering niet alleen een kwestie is van organiseren. Digitale informatie bevat immers een ingebouwd, autonoom mechanisme voor zelfdestructie. De voortschrijdende ontwikkeling van de informatietechnologie leidt tot snelle technologische veroudering. Hoe goed we informatie ook bewaren, zonder verdergaande maatregelen gaat die toch verloren. Het gevolg van een en ander is geheugenverlies van onze samenleving. Als we daar niets aan doen, is in de informatiemaatschappij de toekomst van het heden uiterst onzeker.<sup>3</sup>

De problematiek die ik hier heb geschetst, behoort tot het terrein van wat we tegenwoordig ‘digitale archivering’ noemen.<sup>4</sup> Omdat dat een van de terreinen is waar we vanuit het vakgebied informatiewetenschap momenteel naar kijken, heb ik het als voorbeeld genomen om u duidelijk te maken dat de informatiewetenschap zich bezighoudt met zaken die niet helemaal onbelangrijk zijn. Ik kom er tegen het einde van mijn betoog op terug.

## **Documentaire informatiewetenschap en communicatiewetenschap**

U verwacht misschien van mij dat ik hier een poging doe om u uit te leggen wat informatiewetenschap is. Aan die verwachting zal ik niet helemaal tegemoet komen. Onze samenleving is immers doortrokken van informatie. Met alleen al het beschrijven van alle terreinen waarop informatie een rol speelt, zou ik de mij gegunde tijd overschrijden, zonder zelfs maar toe te komen aan een definitie van het begrip informatie zelf. Maar gelukkig mag ik een beperking aanbrengen door alleen in te gaan op het terrein van mijn leeropdracht, namelijk de documentaire informatiewetenschap.

Het onderwerp van de *documentaire* informatiewetenschap is de rol van informatie als drager van kennis. Kennis wordt geboren in het hoofd van een enkeling, soms ook (tegenwoordig steeds vaker, lijkt het) in de hoofden van een beperkte groep mensen. Kennis zit bij die mensen ‘tussen de oren’, en het is ‘kennis in enkelvoud’, bij dat ene individu of binnen die ene groep. Waar het nu om gaat, is dat die kennis ook bij anderen ‘tussen de oren’ geraakt: het doel is ‘kennis in veelvoud’.<sup>5</sup> Dat gebeurt al eeuwen door die kennis vast te leggen en te verspreiden in de vorm van documenten.<sup>6</sup>

Op het begrip ‘document’ ga ik hier niet uitgebreid in; ik volsta hier met enkele kenmerken:

- documenten bevatten kennis in de vorm van ‘human-readable’ informatie;
- die informatie is vastgelegd op een informatiedrager en kan dus bewaard worden;
- een document kan verveelvoudigd worden, c.q. de informatie die het document bevat, kan gekopieerd worden naar een andere drager;
- de informatie heeft de vorm van (of kan herleid worden tot) een betogende en gestructureerde tekst; deze tekst kan ook niet-tekstuele elementen (zoals al of niet bewegende beelden) bevatten.

Ook het begrip ‘kennis’ zal ik hier niet uitgebreid bespreken. Ik wijs er alleen op dat de kennis waarvan het document de drager vormt, niet beperkt hoeft te zijn tot ‘wetenschappelijke’ kennis. Waar het om gaat, is de verveelvoudiging van

‘producten van de geest’ (kennis, ideeën, opvattingen en dergelijke) door het verspreiden ervan in de vorm van documentaire informatie. Die ‘producten van de geest’ vat ik in dit betoog maar kortweg samen met het begrip ‘kennis’.

Die kennis is in onze samenleving belangrijk. De rol van kennis is zelfs zo belangrijk geworden, dat vaak wordt gesproken over de ‘kennismaatschappij’ en de ‘knowledge based organization’ als organisatieconcept. Zulke organisaties moeten over kennis kunnen beschikken op terreinen als wetenschappelijke en technologische ontwikkelingen, internationale markten, distributiekkanalen, economische, sociale en demografische ontwikkelingen, overheidsbeleid, juridische zaken - om maar enkele voorbeelden te noemen. Niet alleen moet men kennis hebben over al deze zaken, maar men moet ook weten wat *anderen* ervan weten en denken. In de informatiemaatschappij is de kwaliteit van de kennis waarover organisaties kunnen beschikken niet alleen afhankelijk van het opleidingsniveau van individuen, maar ook - en in toenemende mate - afhankelijk van mogelijkheid om binnen de organisatie elkaars kennis te delen, en om te beschikken over kennis die van buitenaf wordt aangeleverd. Dat leidt dan tot het concept van ‘kennismanagement’; daar kom ik straks nog op terug.

Het uitgangspunt van dit alles is dat mensen een uitgesproken behoefte hebben om hun geestesproducten aan de wereld te communiceren, dat anderen er evenzeer behoefte aan hebben om er kennis van te nemen, en dat het om allerlei redenen maatschappelijk van belang is dat deze verveelvoudiging van kennis tot stand komt. Documentaire informatie is een van de middelen die wij daarvoor kunnen gebruiken. De documentaire informatiewetenschap bestudeert de aard van dat communicatiemiddel, de manier waarop mensen er gebruik van maken, de inrichting van het communicatieproces, de daarbij betrokken instituties, de toegepaste netwerken en systemen, en dergelijke. Het zal u dan ook niet verbazen dat ik de documentaire informatiewetenschap beschouw als subdiscipline van de communicatiewetenschap.<sup>7</sup> Het zal u ook duidelijk zijn dat ik niet pleit voor een documentaire informatiewetenschap die zich beperkt tot de studie van het verschijnsel documentaire informatie ‘an sich’, maar voor de studie van communicatieprocessen waarin documenten (met de hierboven omschreven kenmerken) als medium fungeren.

## **De informatieketen**

Een van de bekendere metaforen waarmee het domein van de informatiewetenschap in beeld wordt gebracht, is die van de informatieketen.<sup>8</sup> Deze keten beschrijft het traject dat kennis in documentaire vorm aflegt van de auteur naar de ‘eindgebruiker’. In zijn meest eenvoudige vorm vinden we tussen deze twee polen een aantal instituties, ieder met hun eigen functie. In de klassieke informatieketen zijn dit:

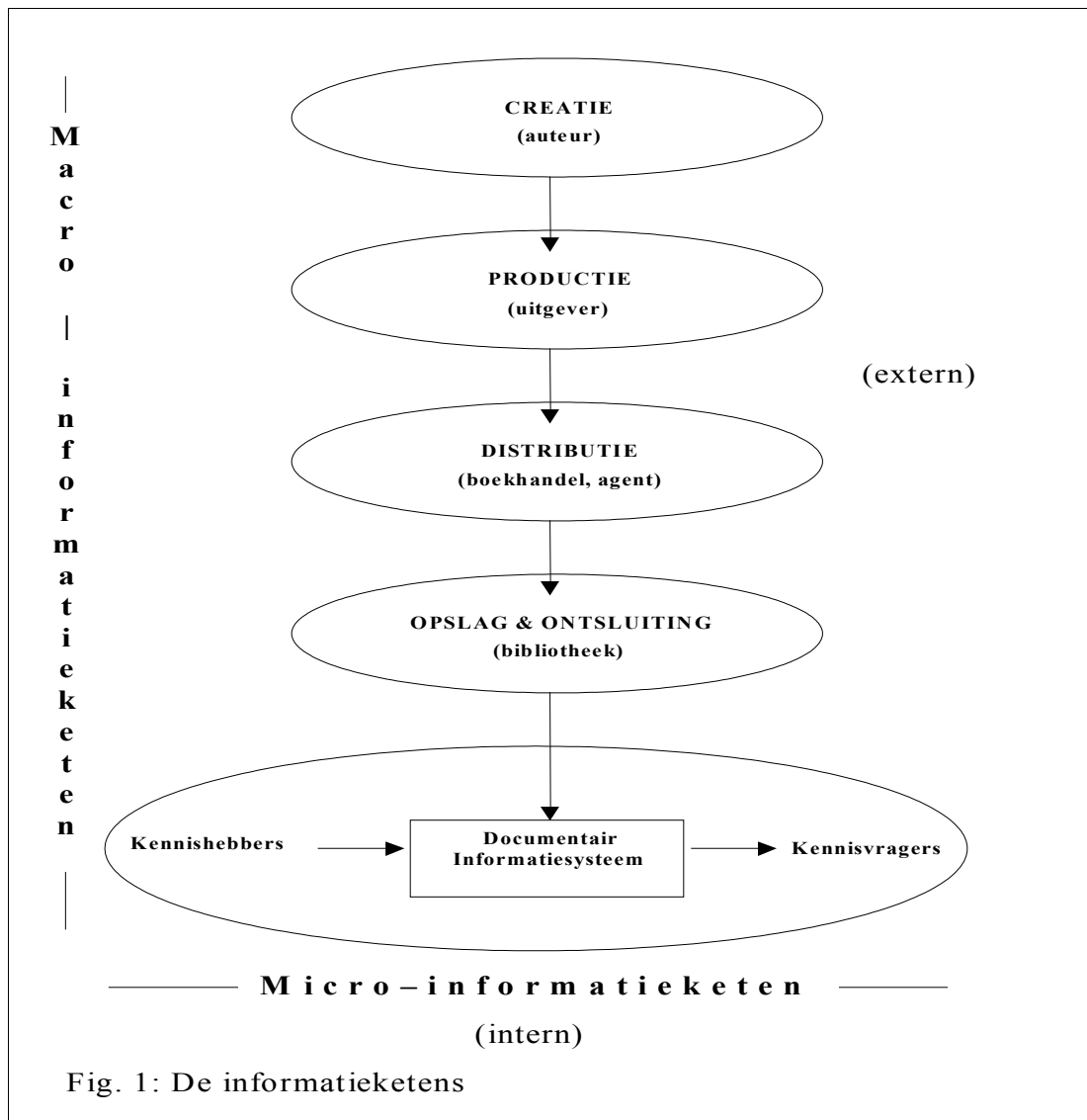
- uitgevers, die de door een auteur aangeleverde informatie in de vorm van een informatieproduct op de markt brengen;
- boekhandels en tijdschriftagenten (‘subscription agents’), die verkoop en distributie voor hun rekening nemen;
- intermediaire organisaties zoals bibliotheken, die verantwoordelijk zijn voor het selecteren van informatie (uit het marktaanbod) en voor het toegankelijk maken en beschikbaar stellen van die informatie ten behoeve van de eindgebruiker.

Er zijn op dit model tal van variaties mogelijk, en er kunnen allerlei andere institutionele partijen een rol in spelen. Verder is het duidelijk dat de inrichting van

de informatieketen door de huidige technologische ontwikkelingen sterk aan het veranderen is. Ook hier kom ik nog op terug.

Deze informatieketen vormt het terrein van wat door sommigen wel wordt aangeduid als de 'informatie-industrie': de economische sector die zich bezighoudt met het produceren en distribueren van 'kennis in veelvoud'. Spelers in dit veld zijn organisaties die zich specifiek bezighouden met het produceren, distribueren, verkopen, organiseren en leveren van informatie. Daartoe behoren zoals gezegd uitgevers, boekhandels en bibliotheken, maar ook databankorganisaties, multimediatelevisieproducenten, 'information brokers', en - steeds meer - adviesbureaus, financiële instellingen en overheden. De informatieketen is in deze vorm een typisch macroverschijnsel; het gaat om een substantiële economische sector en we kunnen heel goed spreken over de mondiale informatieketen.

Naast deze 'externe' informatieketen op macroniveau bestaat er ook een 'interne' informatieketen op microniveau (zie fig. 1). Want binnen iedere kennisintensieve organisatie vinden we een communicatieproces waarbij kennis (in brede zin) in de vorm van informatie wordt vastgelegd, gedistribueerd, beheerd en benut, waarin documentaire informatie dus een communicatieve functie vervult. Ook hier is er sprake van institutionalisering, in de vorm van afdelingen met specifieke verantwoordelijkheden voor onderdelen van het communicatieproces. Ook hier wordt gebruikgemaakt van verschillende media, netwerken en systemen. Ook hier brengen technologische ontwikkelingen grote veranderingen teweeg, met nieuwe mogelijkheden en nieuwe problemen. Deze informatieketen op het microniveau, het communicatieproces binnen organisaties waarbij er sprake is van kennisoverdracht door middel van documentaire informatie, heeft binnen de informatiewetenschap minder aandacht gekregen dan de informatieketen op het macroniveau.<sup>9</sup> Langzamerhand wordt echter duidelijk dat het succesvol functioneren van kennisintensieve organisaties mede afhankelijk is van het functioneren van de interne informatieketen.



## Opvattingen over documentaire informatiewetenschap

Door het domein van de documentaire informatiewetenschap te beschrijven in termen van de beide informatieketens, wek ik bij u de indruk dat het in dit vakgebied vooral gaat om het bestuderen van de inrichting en het functioneren van documentaire communicatieprocessen, zowel in de samenleving als geheel als binnen specifieke organisaties. Dat is ook de manier waarop ik mij voorstel invulling te geven aan het vak. Maar er bestaan ook wel andere opvattingen over informatiewetenschap. Ik wil er enkele voorbeelden van geven, omdat ik meen dat ze op zich nuttig zijn, en ook kunnen bijdragen aan de meer algemene en communicatiewetenschappelijke invulling van de documentaire informatiewetenschap.<sup>10</sup>

### *Informatiewetenschap als literaturenwetenschap*

Een treffende opvatting is die van informatiewetenschappers die hun vak opvatten als 'literatuurwetenschap'.<sup>11</sup> Daarmee bedoelen zij: de wetenschap die zich bezighoudt met het fenomeen van de 'wetenschappelijke literatuur', of zelfs (in de

vorm van 'literaturenwetenschap') van de afzonderlijke literaturen van verschillende disciplinaire domeinen. De nadruk ligt hier op de 'bovenkant' van de informatieketen. Door het bestuderen van de documentaire productie van auteurs, en van hun onderlinge relaties (co-auteurschap, citatie en co-citatie, onderwerpsanalyse) wordt de structuur van wetenschappelijke disciplines blootgelegd en ontstaat duidelijkheid over de ontwikkeling van wetenschappelijke disciplines in de tijd.

Deze vorm van informatiewetenschap, onder meer gerepresenteerd door het terrein van de bibliometrie (tegenwoordig vaak aangeduid met de juistere term 'scientometrie'), behoort primair tot het domein van de wetenschapssociologie.<sup>12</sup> Onderzoek op dit gebied draagt echter ook bij aan het inzicht rond vraagstukken van selectie en navigatie in de 'informatieruimte', en daarmee aan de kwaliteit van de informatievoorziening.

### *Information retrieval*

Bij een tweede benadering van informatiewetenschap, door sommigen beschouwd als de kern van het vakgebied, wil ik wat uitgebreider stilstaan. Bij deze benadering gaat de aandacht vooral uit naar het documentaire 'systeem' dat het mogelijk maakt om uit het beschikbare aanbod aan documentaire informatie juist die informatie te selecteren die de op dat moment gevraagde kennis overdraagt.<sup>13</sup>

Deze benadering is gebaseerd op het zogenaamde IR-paradigma, waarbij 'IR' staat voor 'Information Retrieval'.<sup>14</sup> De IR-benadering ziet, kort samengevat, als kernvraagstuk hoe een systeem kan worden ontwikkeld dat in staat is om uit een gegeven informatieverzameling precies die informatie te selecteren die voldoet aan een gegeven vraag.<sup>15</sup> Meer specifiek moeten we hierbij denken aan een verzameling documenten (of representanten van documenten, d.w.z. beschrijvingen die naar documenten verwijzen) en aan vragen die bestaan uit zoektermen (of een andere formulering) waarmee een gebruiker de gezochte informatie karakteriseert.

Information retrieval is noodzakelijk als gevolg van een drietal anomalieën in de informatieketen: er is te veel informatie, de informatie heeft geen specifieke bestemming, en de informatie blijft liggen tot iemand het vindt. De eerste anomalie heeft betrekking op het bekende verschijnsel van de 'information overload'. De tweede heeft niet zozeer te maken met het feit dat de auteur zijn lezers niet kent, maar met het feit dat er geen eenduidige relatie hoeft te bestaan tussen de inhoud van een publicatie en de informatiebehoefte van de gebruiker. De derde anomalie ontstaat doordat documentaire informatieketens doorgaans over een geheugenfunctie beschikken, bijvoorbeeld in de vorm van bibliotheekcollecties en tekstdatabanken.<sup>16</sup> Zoals altijd bij grote verzamelingen heterogeen materiaal is enige ordening noodzakelijk om chaos te voorkomen.

Onderzoek op het terrein van de information retrieval spitst zich toe op de volgende aspecten:

- de representatie van de inhoud van documenten (de informatiebronnen);
- de representatie van de gezochte informatie (de zoekvragen);
- het mechanisme waarmee selecties worden uitgevoerd door middel van een matching tussen bron- en vraagrepresentatie;
- de criteria voor beoordeling van de kwaliteit van resultaten van het selectieproces (zoals 'recall', 'precision' en 'utility');
- de interactie tussen gebruiker en systeem.

In de loop der jaren is er een bonte verscheidenheid aan technieken ontstaan voor de ogenschijnlijk eenvoudige taak om relevante documenten uit een documentverzameling te selecteren.. Hiertoe behoren onder meer:

- logische operaties op aan de documentverzameling ontleende termen of concepten;
- het gebruik van classificaties, woordsystemen en ontologieën;
- statistische technieken en ‘relevance ranking’;
- linguïstische technieken gericht op syntactische en semantische analyse en het gebruik van lexica;<sup>17</sup>
- technieken ontleend aan het terrein van de kennistechnologie;<sup>18</sup>
- lerende systemen op basis van feedback van de gebruiker.<sup>19</sup>

De ontwikkeling van de information retrieval gedurende de afgelopen decennia is vooral een ontwikkeling geweest in de richting van automatische full-text-indexering. Het handmatig ontsluiten van documenten, bijvoorbeeld door het toekennen van trefwoorden en/of classificaties, is immers gegeven de toenemende hoeveelheden informatie tijdrovend en kostbaar. Juist op dit onderdeel van het vakgebied information retrieval, de ‘automatic indexing’, is veel onderzoek verricht en zijn belangrijke resultaten behaald. Daardoor zijn de kosten van het gebruik van grote documentverzamelingen aanzienlijk gedaald. Het beste voorbeeld hiervan is die megaverzameling van documentaire informatie: het Internet. Zonder de technieken die in de afgelopen decennia op het gebied van automatische indexering en information retrieval zijn ontwikkeld, zou de bruikbaarheid van het Internet als vindplaats van informatie vrijwel nihil zijn. Het is overigens opvallend dat van alle beschikbare technieken er slechts een klein deel ten behoeve van ‘search engines’ op het Internet wordt toegepast.<sup>20</sup> Er is op terrein dan ook nog heel veel mogelijk.

Het onderzoek op dit gebied heeft dus wel geleid tot technieken - en daarop gebaseerde systemen - die het zoeken in grote documentverzamelingen redelijk goed mogelijk maken. Redelijk wil hier zeggen: wel bruikbaar, maar zeker niet volmaakt. In ieder geval zeer aanzienlijk goedkoper dan met handmatige technieken, maar toch niet per definitie zoveel beter.

In het algemeen blijkt dat information retrieval beter lukt binnen een beperkt onderwerpsgebied met een eenduidige en gestandaardiseerde terminologie (zoals de geneeskunde) dan op een diffuus terrein met een ambiguë terminologie (zoals het openbaar bestuur). De laatste jaren worden op het gebied van de information retrieval nog altijd verbeteringen bereikt.<sup>21</sup> Vooral combinaties van technieken, gebaseerd op de verschillende uitgangspunten, blijken effectief. Intussen is wel duidelijk geworden dat ambiguïteit van de taal, zowel die van auteurs als die van informatiezoekers, de grootste belemmering vormt om relevante informatie te scheiden van niet-relevante informatie op basis van een zoekvraag.<sup>22</sup>

In de marge van dit overzicht wil ik graag een kanttekening plaatsen. De technieken die wij momenteel in de context van information retrieval gebruiken, zijn vrijwel allemaal gebaseerd op het a priori maken van een representatie van de inhoud van een document. Aan de ‘retrieval’, het opzoeken van relevante informatie, gaat steeds ‘indexing’ vooraf: het ontlelen aan het document van inhoudomschrijvende termen of concepten<sup>23</sup> en aanvullende gegevens zoals frequenties, relatieve posities, syntactische en semantische relaties, en dergelijke.



Pas als de betekenis van alle documenten letterlijk ‘vaststaat’, is het systeem gereed voor gebruik.<sup>24</sup>

Deze werkwijze gaat voorbij aan iets heel wezenlijks, namelijk dat de betekenis van een document helemaal niet vaststaat. De betekenis van het document dat u nu leest, kan voor u heel anders zijn dan voor uw buurman, kan over tien jaar voor u heel anders zijn dan nu, en kan ook een heel andere zijn dan wat de auteur heeft beoogd. Meestal is men er immers niet op uit om kennis te nemen van het document, maar om het antwoord te vinden op een vraag, of een oplossing van een probleem. Het is daarom heel goed denkbaar dat twee individuen die dezelfde zoekvraag stellen aan hetzelfde systeem, heel anders oordelen over het geleverde resultaat, dat wil zeggen over de relevantie van de gevonden documenten en over ‘recall’ en ‘precision’ van het systeem.

Er is hier iets fundamenteels aan de orde dat bij onderzoekers op het gebied van information retrieval onvoldoende wordt onderkend. Bij de huidige technieken en systemen wordt uitgegaan van het aloude ‘book-on-hand’ principe. Hierbij geldt dat de betekenis van een document volledig en uitsluitend kan worden afgeleid van het document zelf, en begrepen moet worden op basis van het vocabulaire c.q. de ontologie van het domein waarin het document is ontstaan. Binnen deze opvatting moet weliswaar worden erkend dat de wereld nu eenmaal niet volmaakt is. Ambigüiteit, bijvoorbeeld, maakt de betekenis van een document, zeker op het microniveau van woorden en zinnen - vatbaar voor verschillende interpretaties. Maar deze beperking doet geen afbreuk aan het principe dat het document de enige wereld is vanwaaruit de betekenis van het document kan worden afgeleid, ook wanneer dat niet volmaakt mogelijk is.

Eenvoudig gezegd komt dit neer op het nobele streven na te gaan (en te representeren) wat de *auteur* heeft bedoeld. Het komt me echter voor dat de betekenis van een document mede wordt bepaald door het doel waarvoor het document wordt gebruikt, c.q. de kennisbehoefte op grond waarvan het document wordt gezocht. In de praktijk heeft het daarom meer zin om na te gaan wat de *gebruiker* beoogt, en het document te interpreteren op basis van zijn of haar vocabulaire en wereldbeeld.<sup>25</sup> De analyse en representatie van de betekenis van een document moet daarom steeds contextueel zijn, letterlijk ‘in termen van’ de organisatie, het individu of zelfs de individuele vraagstelling. Dit betekent dat (automatische) indexering gebaseerd moet zijn op kennis van de organisatie, en van de probleemcontext van het individu op het moment dat de zoekvraag wordt gesteld.<sup>26</sup> Dit stelt twee voorwaarden aan het retrievalstelsel, namelijk dat het in staat is zich een beeld te vormen van de probleemcontext van de gebruiker, en dat het in staat is om, in ‘real time’, de betekenis van documenten te analyseren in relatie tot die probleemcontext. In deze, door sommigen wellicht als relativistisch en ‘postmodern’ beschouwde richting is er voor onderzoek op het gebied van information retrieval nog veel te doen.<sup>27</sup>

### *De cognitieve benadering*

Tot zo ver een aantal opmerkingen over information retrieval. Ik vervolg mijn bespreking van visies op informatiewetenschap met de zogenaamde cognitieve benadering<sup>28</sup> die gebaseerd is op de gedachte dat informatievoorziening leidt tot een verandering in de ‘kennistoestand’ van een systeem, organisatie of individu. Hiervoor is door Brookes<sup>29</sup> de volgende ‘fundamental equation’ opgesteld:

$$K[S] + \Delta I = K[S + \Delta S]$$

Hiermee wordt aangegeven dat het toevoegen van informatie  $\Delta I$  aan een gegeven 'kennistoestand'  $K[S]$  leidt tot een nieuwe kennistoestand die in  $\Delta S$  van de gegeven kennistoestand afwijkt. Brookes beschouwde de *interpretatie* van deze formule als de belangrijkste opgave van de informatiewetenschap.

De cognitieve benadering is breder dan de IR-benadering, omdat ook wordt gekeken naar de invloed van de gebruiker op het retrievalproces. Het selecteren van relevante kennis wordt niet alleen gezien als een objectief proces van 'matching' van een gegeven vraag op een verzameling beschikbare kennisobjecten. Veeleer gaat het om de houding van het individu ten opzichte van kennis, zijn of haar zoekgedrag, wijze van vraagformulering en manier van interacteren met het systeem, en de daarbij behorende subjectieve, individueel bepaalde karakteristieken.

Deze benadering is tot op heden problematisch gebleken en heeft tot nog toe tot minder toepasbare resultaten geleid dan de 'klassieke' information retrieval. De oorzaak daarvan is ongetwijfeld gelegen in de moeilijkheid om een begrip als 'de kennistoestand van een individu' te operationaliseren.<sup>30</sup> Ik meen echter dat verder stappen in deze richting - op basis van een combinatie van gedragswetenschappelijk en kennistechnologisch onderzoek - buitengewoon vruchtbaar kunnen zijn.

#### *De 'human approach to information science'*

Een minder abstracte benadering vanuit de hoek van de 'cognitive approach' is de zogenaamde 'human approach', die uitgaat van de interactie tussen producenten en consumenten van documentaire informatie.<sup>31</sup> Deze benadering sluit goed aan bij mijn opvatting over informatiewetenschap als communicatiewetenschap. Het begrip documentaire informatie verwijst hier naar die vorm van interactie waarbij vastlegging en opslag van informatie een kenmerk is van het interactiesysteem. In die zin praten we over hetzelfde domein als bij het IR-paradigma. Centraal bij deze benadering staat echter niet het zoekstelsel, maar de wijze waarop de overdracht van producent naar consument wordt georganiseerd. De IR-benadering houdt zich bezig met de kwaliteit van het selectiesysteem. Bij de 'human approach' gaat het om de afstemming tussen vraag en aanbod, en om de toegankelijkheid van informatie. Met andere woorden: hoe kan de informatievoorziening zodanig worden *georganiseerd* dat de informatie die een individu wenst te bezitten ook voor hem of haar beschikbaar is. Het accent ligt hier dus niet op documentaire systemen, maar op organisatievraagstukken met betrekking tot de documentaire informatievoorziening.

Een interessante definitie die deze benadering weergeeft, is de volgende:

...information science considers communication of, mainly recorded, information between human generators and human users. It aims at conceptualizing and understanding information environments, information needs of, and seeking by, individuals and groups, and the organization of information resources for facilitating access to desired information.<sup>32</sup>

Hierin kunnen we een aantal elementen onderscheiden die karakteristiek zijn voor deze benadering:

- Volgens deze definitie heeft het vakgebied niet zozeer betrekking op het verschijnsel informatie als zodanig (wat is informatie eigenlijk, hoe komt het tot stand, welke eigenschappen bezit het), maar op handelingen met betrekking

tot informatie in een maatschappelijke of organisatorische context.

- Informatiewetenschap is primair de wetenschap van informatieoverdracht. Een verdere - en cruciale - inperking van het domein wordt gegeven door de toevoeging 'mainly recorded'. Het vakgebied spitst zich dus toe op de overdracht van een bepaald type informatie. Wij gebruiken hiervoor de term 'documentaire informatie'.
- De definitie geeft aan informatiewetenschap een sterk instrumentele dimensie. Het gaat niet om de empirische vraag hoe informatie wordt overgedragen, maar om de vraag hoe de informatieoverdracht zodanig kan plaatsvinden dat de gewenste informatie optimaal toegankelijk is. Dit kan worden opgevat als de fundamentele vraag of doelstelling van de informatiewetenschap.
- Niettemin verwijst de definitie naar drie empirische gebieden die wezenlijk zijn voor beantwoording van de fundamentele vraag. Deze zijn: de 'information environment' (dat wil zeggen de beschikbare informatie), de gebruikers van informatie (hun informatiebehoeften en zoekgedrag) en de wijze waarop de informatievoorziening reeds is georganiseerd (systemen, infrastructuur).
- Om aan haar doelstelling te voldoen, moet de informatiewetenschap deze drie gebieden niet alleen onderzoeken (onderzoek), maar ook uitspraken doen over veranderingen die tot verbetering van de toegankelijkheid van informatie kunnen leiden (ontwikkeling).

Informatiewetenschap heeft derhalve volgens deze benadering als doel om te onderzoeken welke veranderingen in het aanbod, de organisatie en het gebruik van documentaire informatie kunnen leiden tot optimalisering van de informatievoorziening. Deze benadering is enerzijds breed en omvat de hele 'informatieketen' van producent tot consument. Ook is deze benadering goed toepasbaar op beide door mij onderscheiden informatieketens, dat wil zeggen zowel op macro- als op microniveau. Anderzijds is deze benadering door haar instrumentele karakter ook weer beperkt: aspecten van documentaire informatie die niet bijdragen aan de kwaliteit van de informatievoorziening vallen er buiten.

### *Information engineering*

Een nog duidelijker toegepaste variant van informatiewetenschap staat bekend onder de naam 'information engineering'.<sup>33</sup> Hier staat de ontwikkeling, productie en marketing van (elektronische) informatieproducten, -diensten en -systemen centraal. De nadruk ligt op informatie- en communicatietechnologie, d.w.z. op multimedia, netwerken, hypertext, SGML, grafisch ontwerp, ontwikkelmethodieken en dergelijke. Onlangs is door een Europese projectgroep een curriculum Information Engineering ontwikkeld dat waarschijnlijk in een aantal landen tot een opleiding op dit terrein zal leiden. In Nederland bestaan plannen om dit curriculum op HBO-niveau in te voeren. Information engineering is om twee redenen interessant. Ten eerste leidt het tot inzicht in de aard van informatie- en communicatietechnologieën en nieuwe media die voor informatieoverdracht benut kunnen worden. Ten tweede verschaft het inzicht in de effecten van telematica (informatiedistributie via digitale netwerken) op documentaire communicatieprocessen.

### *Kennismanagement*

Het denken over de organisatie van de interne informatievoorziening (dus op het microniveau) krijgt sinds enige tijd een impuls door de opkomst van het begrip 'kennismanagement'. Kennismanagement is een aanpak die de rol van kennis in en voor de organisatie centraal stelt.<sup>34</sup> Kennismanagement heeft als uitgangspunt dat kennis en kenniswerk kunnen worden geïdentificeerd en in kaart gebracht, dat er procedures voor het genereren, beheren, distribueren en toepassen van kennis kunnen worden ontwikkeld, en dat kennis daartoe kan worden geëxpliciteerd en vastgelegd in informatiesystemen. Tal van ontwikkelingen op het gebied van de informatie- en communicatietechnologie (ICT), zoals netwerken, documentaire informatiesystemen, datawarehousing en datamining, groupware en 'computer supported collaborative work', 'expert systems' en dergelijke, bieden nieuwe mogelijkheden om de 'kennishuishouding' van de organisatie effectief te managen.

Waar het bij kennismanagement onder meer om gaat, is dat we bij het organiseren van de informatievoorziening verder kijken dan naar administratieve gegevens. Bij kenniswerk zijn niet alleen feitelijke gegevens over de interne bedrijfsvoering van belang, maar ook inhoudelijke informatie over procedures, lopende en afgesloten projecten, methoden en technieken, externe relaties, ontwikkelingen in de markt, technologie, concurrenten, economische, sociale en juridische ontwikkelingen, overheidsbeleid en dergelijke. De procedures en systemen die in het kader van kennismanagement worden ontwikkeld, dienen er dan ook voor om kenniswerk te ondersteunen: het werk van bijvoorbeeld managers, stafmedewerkers, beleidsontwikkelaars, onderzoekers en adviseurs.

Door deze oriëntatie op inhoudelijke informatie die zich goed leent voor vastlegging en distributie in de vorm van tekstuele/documentaire informatie, kan de informatiewetenschap uitstekend als hulpwetenschap ten behoeve van kennismanagement dienen. Het gaat bij kennismanagement immers voor een belangrijk deel om hoogwaardige communicatie: het uitwisselen van kennis tussen en het overdragen van kennis aan mensen in de organisatie. Een van de manieren om dat te bereiken, is kennisuitwisseling en kennisoverdracht 'in het werk', bijvoorbeeld binnen projectteams waar men elkaars kennis opdoet door samen te werken en van elkaar te leren. Maar een andere manier is het vastleggen en verspreiden van kennis in documentaire vorm. De technieken en systemen die worden ontwikkeld vanuit de information retrieval, maar vooral de methoden en inzichten die worden ontwikkeld vanuit de 'human approach', zijn hier uitermate bruikbaar.

Kennismanagement vormt ook bij uitstek de context voor onderzoek naar de informatieketen op microniveau, naar de documentaire communicatieprocessen binnen organisaties. Daarnaast is er natuurlijk ook sprake van kennisuitwisseling tussen de organisatie en zijn omgeving. Organisaties dragen hun kennis naar buiten uit, en zijn mede afhankelijk van kennis over de buitenwereld. Ook hier, op het koppelpunt tussen het microniveau en het macroniveau van de informatieketen, ligt een interessant terrein voor de informatiewetenschap.

Ik hecht eraan te benadrukken dat informatiewetenschap en kennismanagement twee verschillende disciplines zijn. Professionals op het gebied van de *documentaire* informatievoorziening<sup>35</sup> hebben een sterke neiging om kennismanagement voor zich op te eisen en te vereenzelvigen met hun werkterrein. Hetzelfde zien we bij IT-specialisten. Kennismanagement is echter geen specifieke bedrijfsfunctie en kan niet worden toegewezen aan de ene of de andere afdeling van een organisatie. Kennismanagement is de verantwoordelijkheid van het topmanagement en vormt een geheel van activiteiten, methoden en opvattingen die in alle bedrijfsfuncties en op alle niveaus in de organisatie aan de orde is. De documentaire informatievoorziening - met als basis de documentaire informatiewetenschap - kan (en moet) net als de IT-functie, Human Resources

Management en andere vakgebieden aan kennismanagement bijdragen. Kennismanagement behoort echter aan geen van deze gebieden toe, maar aan de organisatie als geheel.<sup>36</sup>

### *Een pragmatische benadering*

Om te komen tot een vruchtbare documentaire informatiewetenschap is mijns inziens een synthese nodig van de verschillende benaderingen die ik hiervoor heb beschreven. Ik zie dat vooral in de context van een pragmatische benadering op basis van het gegeven dat documentaire informatie nu eenmaal als drager van kennis een functie vervult in de samenleving. Er bestaat, met andere woorden, een maatschappelijke activiteit die we hier aanduiden met het begrip ‘documentaire informatievoorziening’. Daarin doen zich tal van vraagstukken voor, waarvoor vanuit de samenleving (op macro- én op microniveau) oplossingen worden gevraagd. Deze vraagstukken zijn relevant voor verschillende actoren die op velerlei manieren belang hebben bij de informatievoorziening. Dat kan zijn als gebruiker (waar de ‘human approach’ zich vooral op richt), als organisatie (het aandachtsgebied van kennismanagement), als professionele of commerciële belanghebbende, vanuit een economisch, sociaal of cultureel belang, of anderszins. Het is de taak van de informatiewetenschap om op vraagstukken met betrekking tot de documentaire informatievoorziening adequate, bevredigende antwoorden te geven.

De term ‘pragmatisch’ verwijst naar het feit dat de informatiewetenschap niet gebaseerd is op een theoretisch concept of paradigma, maar op een maatschappelijk verschijnsel. De grenzen van de informatiewetenschap zijn niet a priori gegeven, maar worden bepaald door behoeften vanuit de samenleving. De problemen die zich in de praktijk van de informatievoorziening op een gegeven moment voordoen, zijn per definitie relevant voor de informatiewetenschap en bepalen er de inhoud van. Dit laat uiteraard onverlet dat binnen de informatiewetenschap theorieën kunnen worden ontwikkeld die bijdragen aan het oplossen van vraagstukken in de praktijk. De verschillende theoretische invalshoeken (zoals het IR-paradigma, de cognitieve benadering, de bibliometrische benadering, de ‘human approach’ en andere) die aan de oplossing van dergelijke vraagstukken kunnen bijdragen, vormen in mijn opvatting het instrumentarium van de informatiewetenschap. In deze opvatting is de informatiewetenschapper onderzoeker, ontwerper, architect en ingenieur.

### **Ontwikkelingen in de externe informatieketen**

Tot zover deze wat impressionistische bespreking van het terrein van mijn leeropdracht: de documentaire informatiewetenschap. Ik wil nu graag kort ingaan op een aantal ontwikkelingen in de externe informatieketen, omdat dat mij de gelegenheid geeft om aan het einde van deze voordracht terug te keren tot het begin: de vraag in welke mate de informatie van nu in de toekomst nog beschikbaar zal zijn.

Om die vraag te beantwoorden, moeten we eerst kijken naar de huidige ontwikkelingen in de informatieketen. Ik beperk me vandaag tot de ‘macroketen’, waarin, naast auteurs en lezers, met name uitgevers en bibliotheken tot nog toe de dominante spelers zijn geweest.

Er zijn twee technologische ontwikkelingen die momenteel leiden tot ingrijpende veranderingen in de informatieketen: digitalisering en netwerken. Digitalisering houdt in dat informatieproducten die voorheen in gedrukte vorm via de informatieketen werden gedistribueerd, nu in digitale vorm beschikbaar komen.

Netwerken maken het mogelijk dat digitale informatieproducten ook op digitale wijze worden verspreid. Het gehele communicatieproces, vanaf het schrijven door de auteur tot en met het lezen door de eindgebruiker, kan daardoor volledig digitaal plaatsvinden. De traditionele partijen in de informatieketen reageren hier op allerlei manieren op.

#### *Auteurs*

Steeds meer auteurs gaan ertoe over om door hen vervaardigde documenten zelf via het Internet beschikbaar te stellen. Soms is dat de enige vorm van openbaarmaking, soms gaat het vooraf aan de officiële publicatie door een uitgever. In sommige disciplines, zoals de hoge-energiefysica en de cognitieve wetenschappen is de distributie van preprints via het Internet volledig geïnstitutionaliseerd. Steeds meer auteurs gaan ertoe over een homepage op het World Wide Web in te richten waar ze hun eigen publicaties in digitale vorm beschikbaar stellen.<sup>37</sup> Inmiddels zijn door groepen wetenschappers digitale wetenschappelijke tijdschriften opgezet.<sup>38</sup> Deze en andere vormen van 'self-publishing' door auteurs leiden intussen tot heftige discussies over de rol - en zelfs het bestaansrecht - van de wetenschappelijke uitgeverij.<sup>39</sup>

#### *Uitgevers*

Uitgevers hebben betrekkelijk weinig pogingen gedaan om nieuwe digitale informatieproducten te ontwikkelen, en zich vooral geconcentreerd op het digitaliseren van bestaande producten.<sup>40</sup> Met name de grote wetenschappelijke uitgevers beschikken nu over digitale versies van hun tijdschriften. Deze worden in deze vorm, doorgaans tegen een meerprijs, aan bibliotheken geleverd. Er tekent zich echter een trend af waarbij de uitgevers de digitale versies van hun tijdschriften via hun eigen servers op het Internet beschikbaar stellen - uiteraard tegen betaling, doorgaans op basis van licenties.<sup>41</sup> Het valt te verwachten dat deze 'digital repositories' in de toekomst de exclusieve bron voor wetenschappelijke tijdschriften zullen zijn.

#### *Bibliotheken*

Bibliotheken lopen in veel opzichten voorop bij de toepassing van informatie- en communicatietechnologie. Zij zijn zich in het algemeen aan het instellen op de digitale informatieketen op een aantal manieren. Zij dragen bij aan de bewegwijzering van het Internet, treden op als 'contractmanager' bij het afsluiten van licenties die eindgebruikers toegang geven tot digitaal materiaal van uitgevers<sup>42</sup>, en zij richten zich sterk op het ontwikkelen van diensten op de werkplek van de gebruiker. Dit alles leidt op den duur tot bibliotheken die digitaal zijn (want gebaseerd op digitale informatiebronnen)<sup>43</sup>, virtueel (dat wil zeggen als functie geïntegreerd op de 'desktop' van de gebruiker) en gedistribueerd. Met dit laatste bedoel ik dat verschillende, geografisch verspreide bibliotheken gezamenlijk diensten aanbieden die voor de gebruiker transparant zijn, dat wil zeggen door één organisatie verzorgd lijken te worden.<sup>44</sup>

#### *Gebruikers*

De informatiegebruiker ondertussen, beschouwt het netwerk (dat is momenteel het Internet) steeds meer als de primaire informatiebron. Voor zover traditionele partijen als uitgevers en bibliotheken daarop aanwezig zijn, gemakkelijk toegankelijk zijn, en functioneel een toegevoegde waarde hebben, vormen zij een nuttige aanvulling. Wat van belang is, is dat voor de gebruiker de enige informatie die informatie is die direct (dat wil zeggen waar en wanneer zich de vraag voordoet) via het netwerk

beschikbaar is. Andere informatie blijft onzichtbaar, en bestaat dus in feite niet.

Deze simpele (en ook simplistische) constateringën kunnen aanleiding geven tot uitvoerige beschouwingen, bijvoorbeeld over de vraag of er in een digitale informatieketen nog wel behoefte is aan de traditionele partijen als uitgevers en bibliotheken. Duidelijk is in ieder geval wel dat de positie van alle partijen aan het veranderen is. Was er vroeger sprake van een efficiënte taakverdeling, nu zien we dat de verschillende institutionele partijen binnen de informatieketen allemaal proberen dezelfde functies naar zich toe te trekken. Kort gezegd: uitgevers vervullen bibliotheekfuncties en omgekeerd. Ondertussen is allesbehalve duidelijk of de gebruiker nog behoefte heeft aan die functies, ongeacht door wie ze worden vervuld. Zal in de toekomst nog gelden dat 'a scientific laboratory without a library is like a decorticated cat'?<sup>45</sup> En moeten wij ons niet afvragen of de ontwikkelingen op het gebied van digitalisering en netwerken, in samenhang met toenemende monopolievorming en prijsopdrijving, er misschien op wijzen dat de wetenschappelijke uitgeverij aan het einde van zijn levenscyclus begint te komen?<sup>46</sup> Maar weest u gerust, die onderwerpen laat ik vandaag rusten. Ik wil alleen stilstaan bij de vraag waar al die informatie blijft die in toenemende mate digitaal op het netwerk komt te staan, en of er in de toekomst nog wel sprake zal zijn van 'kennis in veelvoud'.

### **Digitale archivering<sup>47</sup>**

Aan het begin van mijn verhaal heb ik u gezegd dat digitale informatie verdwijnt als niemand de verantwoordelijkheid op zich neemt voor de archivering. Er treedt door de digitalisering immers een systeemverandering op. Bij gedrukte publicaties bestaan er vele exemplaren, verspreid over vele bibliotheken, soms over de hele wereld. In fysieke zin: kennis in veelvoud. Zo'n systeem blijkt resistent tegen brand, overstromingen, oorlog en censuur. In het digitale tijdperk is er slechts sprake van publiceren in enkelvoud. Dat geldt zeker als uitgevers hun publicaties via 'digital repositories' aanbieden, en lokale opslag in bibliotheken niet toestaan. Maar uitgevers zullen die publicaties om commerciële redenen niet *langdurig* archiveren en beschikbaar stellen. De meeste bibliotheken hebben niet de technische en financiële mogelijkheden om dat te doen. De vluchtigheid van onze samenleving dreigt weerspiegeld te worden in een onvermogen om onze uitingen van cultuur en wetenschap voor het nageslacht te behouden. Er moet dus iets georganiseerd worden. Wat het meest voor de hand ligt, is dat een beperkt aantal grote organisaties de verantwoordelijkheid voor de digitale archivering voor de lange termijn op zich nemen. Daartoe lijken mij het meest aangewezen de nationale bibliotheken die, in ieder geval in Europa, al een gevestigde traditie hebben als depotbibliotheken voor de nationale productie van gedrukte publicaties.<sup>48</sup>

Al sinds enige tijd zijn veel nationale bibliotheken bezig om de digitale archivering ter hand te nemen. Een van de belangrijkste problemen die zij daarbij tegenkomen, is niet het verkrijgen van het digitale materiaal. Zelfs zonder juridische dwang blijken uitgevers daar wel aan te willen meewerken.<sup>49</sup> Nee, de problemen komen uit een andere hoek: de snelle technologische veroudering van het materiaal, en de juridische belemmeringen<sup>50</sup> die het moeilijk maken om aan die veroudering het hoofd te bieden. Omdat digitale informatieproducten te kunnen bewaren, moeten er namelijk twee dingen mee gebeuren.

Ten eerste moeten zij 'verdoekt' worden.<sup>51</sup> Daarmee bedoel ik dat zij overgebracht moeten worden op gestandaardiseerde media van hoge kwaliteit, waarmee de integriteit van het oorspronkelijke product - op het niveau van de 'bits

en bytes' - wordt gewaarborgd. Dat is overigens een kopiehandeling die door de huidige wetgeving (nog) niet zonder meer wordt toegestaan.

Ten tweede is het nodig om digitale informatieproducten met een zekere regelmaat te 'migreren' naar een nieuwe 'technologische omgeving', naar nieuwe hardware- en softwaresystemen wanneer de oude systemen in onbruik raken. Dat vereist zonder meer een ingrijpen in het informatieproduct zelf, een verandering van de 'bits en bytes', en zal in veel gevallen onvermijdelijk leiden tot informatieverlies en verlies van integriteit. Jammer, onvermijdelijk en verboden. Want welke auteur of uitgever zal toestaan dat er om technische redenen in zijn of haar publicatie veranderingen worden aangebracht? Maar zal men, liever dan het onvermijdelijke te accepteren, het verdwijnen van zijn publicaties voor lief nemen?

Integriteit is bij dit alles het toverwoord: de publicatie zoals die is verschenen, is onaantastbaar. Maar moeten we dat wel zo zien? Sinds de uitvinding van het schrift, en meer speciaal sinds de uitvinding van de boekdrukkunst, is er een onlosmakelijke relatie tussen inhoud (content) en vormgeving (typografie, lay-out, formaat, kleurgebruik etc.). Drukkers, uitgevers en vormgevers hebben zich tussen de auteur en de lezer geplaatst. De vormgeving ligt als een onverbreekelijke visualisatielaag op de tekst, onverbrekelijker dan de folie op het paspoort van uw Staat der Nederlanden. Soms is die laag van vormgeving de auteur en lezer behulpzaam, soms in het geheel niet. Met de boekdrukkunst en (typo)grafische vormgeving is een vakgebied ontstaan - sommigen zeggen zelfs: een kunstvorm - die een eigen leven leidt en zelfs onderwerp vormt van een academische discipline, namelijk de boekwetenschap.

We zien nu bij digitale media een verbreking van de relatie tussen inhoud en vorm, een losmaking van de laag van visualisatie. Lezers kunnen met behulp van technische middelen (zoals 'browsers') hun eigen vorm(geving) creëren, een eigen wijze van zichtbaar maken van de tekst. De vormgeving, de visuele presentatie van de inhoud is nu niet van de maker, maar van de lezer, niet van de producent maar van de consument. De inhoud is niet meer dan de symbolische representatie van de gedachten van de auteur. De lezer maakt die op zijn eigen manier zichtbaar, afhankelijk van zijn aard of haar stemming van het ogenblik.<sup>52</sup>

De rol die vormgevers, uitgevers en drukkers in de loop der eeuwen hebben opgeëist, wordt steeds meer verlegd enerzijds naar de auteur (als distributeur van teksten) en anderzijds naar de lezer (als vormgever van die teksten). Naast het verlies van de distributierol, verliest de uitgever ook zijn rol als vormgever van teksten. Dat is nog een reden om ons af te vragen of digitale media misschien het einde van de levenscyclus van de uitgeverij inluiden.

Als we in de toekomst willen weten wat auteurs hebben gedacht en geschreven, dan moeten we hun teksten bewaren. Ik bedoel daarmee de tekst zelf, niet de publicatie. Wat uitgevers, drukkers en vormgevers er eventueel van gemaakt hebben, de *vormgegeven* tekst, is in de digitale wereld niet wezenlijk relevant meer. De vormgegeven tekst is hooguit interessant als voorbeeld van een mogelijke visualisatie, maar staat als tijdsgebonden cultuuruiting los van de inhoud van de tekst. Als we later willen weten hoe lezers die teksten ooit gezien hebben, dan hebben we een probleem. Omdat iedere 'reading event' een unieke visualisatie kan zijn, bestaat er niet meer zoiets als *de* vorm. Wat we wel kunnen doen, is de visualisatietechniek bewaren, opdat we in ieder geval weten welke mogelijkheden voor visualisatie er waren in de tijd dat een tekst werd geschreven.

De integriteit van de tekst dient altijd voorop te staan. Waar ik daarom voor wil pleiten, is dat we bij de digitale archivering van publicaties juist *niet* overgaan tot transformaties en migratie, tot aanpassing aan nieuwe technologische systemen. Laten we het bij 'verdoeken' laten, en de 'tekst', de reeks 'bits en bytes' van de



publicatie intact laten en in originele staat bewaren. Wat we vervolgens moeten doen, is systemen ontwikkelen die in staat zijn om, met de techniek van het moment, zo goed mogelijk naar dat zorgvuldig geconserveerde origineel te kijken, en daar zoveel informatie uit te halen als op dat ogenblik mogelijk is. Juist omdat wij niet weten over welke technologische mogelijkheden toekomstige generaties zullen beschikken, hebben wij de plicht om digitale publicaties zoveel mogelijk in hun oorspronkelijke vorm te bewaren. Anders treedt er van generatie op generatie, van migratie op migratie, telkens een beetje informatieverlies op, totdat we van het origineel te weinig overhouden.

### **Een toekomst voor het publiceren?**

Tenslotte nog dit. Als ik nadenk over de problematiek van digitale archivering, dan dringt zich aan mij de vraag op of die problematiek, zoals wij die zien aan het einde van de twintigste eeuw, niet voortvloeit uit projectie van onze huidige opvattingen over wetenschappelijke communicatie.

Wat vooral opvalt is dat het gebruik van digitale media nog helemaal gebaseerd is op metaforen van de traditionele schrijf- en uitgeefpraktijk in de wereld van de gedrukte media. De wetenschappelijke ‘body of knowledge’ is in die wereld - noodgedwongen - gebaseerd op lineaire teksten, die op zich weer onderdeel vormen van een reeks in de tijd: een nieuwe publicatie bouwt voort op voorgaande publicaties. Een andere ordening dan de sequentiële opeenvolging van publicaties kennen we in feite niet. In de verzameling van los, via citaties en hyperlinks, aan elkaar gekoppelde publicaties zit geen formeel logische structuur, en de onderdelen van de kennisverzameling zijn min of meer willekeurig verspreid over bibliotheken en - tegenwoordig steeds vaker - document servers op het Internet.

Dat dit niet altijd als een bezwaar wordt gezien, hangt samen met het feit dat wetenschappers het publiceren maar ten delen beschouwen als iets dat met wetenschappelijke communicatie van doen heeft. De hang naar wetenschappelijke status en het ‘publish or perish’ beleid van de universitaire wereld spelen bijvoorbeeld een belangrijke rol. Ik herhaal bovendien een elders door mij gemaakte opmerking dat wetenschappelijke publicaties maar zeer ten dele bedoeld zijn om wetenschappelijke kennis te over te dragen, omdat ze ook allerlei andere functies vervullen.<sup>53</sup>

In de kinderjaren van het digitale tijdperk die wij nu meemaken, doen wij onze voorouders na door een op publicaties gebaseerde ‘body of knowledge’ en communicatiemodel vrijwel één op één te kopiëren naar digitale media. Maar in het digitale tijdperk kan het ook anders. We hebben in de toekomst de mogelijkheid om de ‘body of knowledge’ niet op te vatten als een sequentiële reeks van ‘loosely coupled’ documenten, maar als een bewust gecreëerde logische structuur waaraan iedere wetenschapper zijn steentje bijdraagt.<sup>54</sup> In zo’n wereld dragen auteurs direct stukjes kennis bij aan een gestructureerde ‘body of knowledge’, zonder gebruik te maken van het medium ‘publicatie’. Als we kijken naar het Internet en zien dat het één gigantische informatiebron is met toch een zekere - en hopelijk toenemende - ordening, dan beseffen we dat het begrip ‘publicatie’ nu al een anachronisme aan het worden is.

Zeer gewaardeerde toehoorders,

ik heb aan het begin van mijn betoog u aan de hand van twee voorbeelden laten zien dat digitalisering van de kennisoverdracht tot problemen kan leiden voor wat betreft

de beschikbaarheid van die kennis voor toekomstige generaties. Verder heb ik gesproken over de informatieketens, en heb ik u erop gewezen dat er binnen die informatieketens grote veranderingen op til zijn. Dat zijn veranderingen met een macro-economische dimensie - omdat ze een grote bedrijfstak als de informatie-industrie betreffen, met een micro-economische dimensie - omdat het gaat om het functioneren van kennisintensieve organisaties, en met een wetenschappelijke dimensie - omdat ze te maken hebben met de manier waarop wij kennis distribueren. Ik ben wat dieper ingegaan op het vakgebied informatiewetenschap, om u duidelijk te maken dat mijn vakgebied wel iets met die zaken van doen heeft. Tenslotte ben ik teruggekomen bij het probleem van de digitale archivering. Ik heb betoogd dat we daarbij aandacht moeten hebben voor de mogelijk afnemende rol van uitgevers en bibliotheken, en de verschuiving van functies naar de uiteinden van de informatieketen: naar auteurs en lezers. Ik heb ook betoogd dat deze verschuiving medebepalend is voor de manier waarop we tegen de archiveringsproblematiek aan moeten kijken, en voor de technische oplossingen die we daarvoor moeten ontwikkelen. Tenslotte heb ik die problematiek weer gerelativeerd door te wijzen op de mogelijkheid dat wij in een meer volwassen stadium van de informatiemaatschappij tot heel nieuwe vormen van wetenschappelijke communicatie zullen geraken, waarbij het produceren en dus ook het archiveren van digitale publicaties als zodanig niet meer aan de orde is.

Dat is heel wat voor een achternamiddag in juni, en toch heb ik maar enkele aspecten van mijn vakgebied kunnen aanstippen. Ik ben de Universiteit van Amsterdam dan ook dankbaar voor de mij geboden gelegenheid mij - al is het maar gedurende een (te) bescheiden twee dagen per week - aan deze tak van wetenschap te wijden. Dat de plaats binnen deze organisatie die de Universiteit voor mij en mijn collega's heeft uitgekozen om het vak te beoefenen, aan dat vak geen recht doet, moet op deze feestelijke middag maar onvermeld blijven.

Waarde collega's,

binnen de Leerstoelgroep Boek-, Archief- en Informatiewetenschap vertegenwoordigen wij een bonte verscheidenheid aan vakgebieden. Die verscheidenheid is eigenlijk te groot, en doet ons en onze studenten geen goed. Niettemin prijs ik me gelukkig met een omgeving waar ik zoveel kwaliteit, integriteit, collegialiteit en - niet in de laatste plaats - werklust aantref. Ik hoop zodanig bij te dragen dat u die eigenschappen ook aan mij zult toedichten.

Dames en heren studenten,

tot u richt ik het laatste woord. U bent hier vanmiddag niet in groten getale aanwezig. Niet uit gebrek aan belangstelling voor wat ik u te vertellen heb, maar omdat er gewoon weinig jonge mensen zijn die begrijpen dat informatiewetenschap een boeiend vak is met grote toekomstmogelijkheden. Dat geeft alle aanleiding om u te koesteren en uw ontluikende enthousiasme voor dit mooie vak te voeden met goed onderwijs. Want vergeet u niet dat het úw toekomst is waarvoor wij hier werken.

Dames en Heren, ik heb gezegd.

## NOTEN

- <sup>1</sup> Mackenzie Owen, J.S. - Bibliotheek, informatie & computer. - Den Haag: NBLC, 1986.
- <sup>2</sup> Mackenzie Owen, J.S. - Organizing for digital archiving: new distribution models in the scientific information chain. - In: Butterworth, I. (ed.) - The impact of electronic publishing on the academic community. - London: Portland Press, 1998 (in press).
- <sup>3</sup> De eerste die wees op de risico's van digitalisering voor het behoud van het intellectuele erfgoed is Peter S. Graham. Zie: Graham, P.S. - Long-term intellectual preservation. - In: Elkington, N.E. (ed.) - Digital imaging technology for preservation: proceedings from an RLG symposium held March 17 and 18, 1994: Cornell University, Ithaca, New York. - Mountain View, Calif.: Research Libraries Group, 1995, p. 41-58. -  
[<http://aultnis.rutgers.edu/texts/dps.html>]. en Graham, P.S. - Intellectual preservation: electronic preservation of the third kind. - Washington: Commission on Preservation and Access, March 1994. - [<http://aultnis.rutgers.edu/texts/cpaintpres.html>].
- <sup>4</sup> Mackenzie Owen, J.S.; Walle, J. v.d. - Deposit collections of electronic publications. - Luxembourg: European Commission, 1996 (EUR 16910 EN).
- Mackenzie Owen, J.S. - Preservation of digital materials for libraries. - Paper presented at the LIBER Annual Conference, University of Malta, May 1996. (gepubliceerd in: Liber quarterly, 6(1996)4, p. 435-451).
  - Task Force on Archiving of Digital Information. - Preserving digital information: final report and recommendations. - Washington: CPA. 1996. -  
[<http://www.rlg.org/ArchTF/>].
- <sup>5</sup> Ik raak hier aan een gecompliceerd onderwerp, namelijk de vraag wat kennis is. Volgens gangbare opvattingen maken we onderscheid tussen 'belief' (opinie, datgene wat men meent te weten) en 'justified belief' (algemeen aanvaarde kennis). Daarnaast onderscheiden we 'tacit' en 'explicit' belief. 'Tacit' belief bevindt zich nog 'tussen de oren' en wordt 'explicit belief' door communicatie. Informatieoverdracht door middel van documenten is een van de manieren waarop dit gebeurt. Informatieoverdracht is ook onderdeel van het proces waardoor wetenschappelijke opinies (belief) tot wetenschappelijke kennis (justified belief) worden. In de praktijk is er naar mijn mening bij documentaire informatie steeds sprake van een (overigens problematische) combinatie van opinies en kennis. Voor het gemak gebruik ik de term 'kennis' als overkoepelend begrip. Ik bedoel er dus mee: zowel datgene wat men meent te weten, als datgene wat als vaststaand wordt aanvaard. Zie over het verschil tussen opiniesystemen en kennissystemen: Vet, P.E. van der; Mars, N.J.I. - De kwaliteit van kennis. - In: Mars, N.J.I (red.) - Informatiewetenschap 1992: wetenschappelijke bijdragen aan de tweede STINFON-conferentie, Enschede, 18 december 1992, blz. 77-88.
- <sup>6</sup> Van der Wolk drukte dit ooit uit als: 'Informatie is kennis in beweging tussen mensen' (Wolk, L.J. van der - Literatuurinformatie. - In: Bibliotheekleven, 53(1968)7/8, blz. 353-363). Zie verder het afscheidscollege 'Documentaire informatie als een sociale wetenschap' (Deventer: Van Loghum Slaterus, 1979) van mijn illustere voorganger Th.P. Loosjes.
- <sup>7</sup> Vgl. de definitie van Van Cuilenburg c.s.: '*Communicatiewetenschap is de wetenschap die zich bezighoudt met de bestudering van informatievoorziening die op een beroeps- en/of bedrijfsmatige basis plaatsvindt aan een algemeen danwel een professioneel ontvangerspubliek*'. Zie: Van Cuilenburg, J.J.; Scholten, O.; Noomen, G.W. - Communicatiewetenschap. - Muiderberg, Coutinho, 1992, blz. 33. De opvatting van informatiewetenschap als de studie van communicatieprocessen is niet nieuw, maar was dat in 1987 kennelijk nog wel toen Brian en Alina Vickery in het voorwoord van hun bekende boek schreven: '*This book is an attempt to present and discuss a scientific understanding of the process of information transfer. This transfer is a human, social activity ... In seeking scientific understanding of the process of information transfer we have had to go considerably outside the subject limits within which 'information science' as an academic subject is normally constrained ... It has become increasingly clear that only by widening its 'knowledge base' can information science establish a solid foundation for future development*' (Vickery, B.; Vickery,

A. - Information science in theory and practice. - London etc.: Bowker-Saur, 1987). In meer specifieke zin, en met name binnen het domein van de 'micro-informatieketen', zijn er sterke overeenkomsten tussen informatiewetenschap en het terrein dat binnen de communicatiewetenschap wordt aangeduid als 'informatiekunde': het geheel van kennis en vaardigheden met betrekking tot het verzamelen, opslaan, bewerken, verwerken, presenteren en distribueren van gegevens voor besluitvormingsprocessen (Van Cuilenburg c.s., op. cit., p. 90). Hooguit kan men zeggen dat de informatiewetenschap zich meer met kennisdragende informatie dan met zuivere gegevens bezighoudt, en dat het toepassingsdomein breder is dan dat van de besluitvorming in engere zin. Met dat laatste bedoel ik dat het toepassingsdomein het geheel van denken en handelen binnen organisaties is voor zover die door informatie (als drager van kennis) worden beïnvloed. Maar de term 'besluitvorming' wordt, denk ik, door de communicatiewetenschappers ook wel zo bedoeld.

8 Zie: Duff, A.S. - Some post-war models of the information chain. - In: Journal of librarianship and information science, 29(1997)4, p. 179-187 en: Mackenzie Owen, J.S.; Halm, J. van - Effecten van technologie op de informatieketen. - Den Haag: Bibliotheekraad, 1987 (in verkorte vorm verschenen als: Innovation in the information chain. - London, New York: Routledge, 1989). In deze laatste studie wordt de informatieketen opgevat als een markt waarin er sprake is van vraag en aanbod. In mijn opvatting is de informatieketen, zeker op macroniveau, een (*massa*)*communicatiemarkt*, die in de terminologie van Van Cuilenburg c.s. zowel kenmerken van een informatiemarkt als van een aandachtsmarkt vertoont. In deze markt is er over het algemeen eerder sprake van overvloed ('information overload') dan schaarste, en zijn het de aanbieders die dingen om de aandacht van de afnemer. Vanwege die overvloed vervullen informatiesystemen in de keten vooral de functie van *selectiesystemen*, al worden ze vaak aangeduid met de term *zoeksysteem* (tegenwoordig ook wel 'search engines'). Niettemin wijst de populariteit van het browsen op het Internet ook op een onstilbare honger naar informatie. Of de overvloed aan informatie in stand wordt gehouden door de aanbodzijde, inclusief individuele allocutiebehoeften, wetenschapspolitieke factoren ('publish or perish') en commerciële doelstellingen van uitgevers, of door een reële behoefte aan de vraagzijde, valt niet gemakkelijk na te gaan. De toenemende tijdsbesteding aan informatiezoeken op het Internet doet wel de vraag rijzen in welke mate mensen een rationele afweging maken tussen kosten en opbrengsten van informatiegebruik.

9 Met name wordt voorbijgegaan aan het feit dat documentaire informatie binnen de interne informatieketen betrekking heeft op alle vormen van kennisdragende informatie, en daardoor op een veel breder scala aan 'kennisobjecten'. Dit impliceert een aanzienlijke uitbreiding van het begrip 'document' en zou kunnen leiden tot heel nieuwe toepassingen van documentaire informatiesystemen.

10 Weinigen hebben zoveel teksten gewijd aan de definitie (en rechtvaardiging) van hun vakgebied als de beoefenaren van 'information science'. Een uitvoerige en kritische analyse is: Schrader, A.M. - The domain of information science: problems in conceptualization and in consensus-building. - In: Information services & use, 6(1986), p. 169-205. Zie verder:

- Buckland, M.K.; Liu, Z. - History of information science. - In: Annual review of information science and technology, 30(1995), p. 385-416.
- Olaisen, J.; Munch-Petersen, E.; Wilson, P. (eds) - Information science: from the development of the discipline to social interaction. - Oslo: Scandinavian University Press, 1996.
- Vakkari, P.; Cronin, B. (eds) - Conceptions of library and information science: historical, empirical and theoretical perspectives. - London: Taylor Graham, 1993.

11 Zie: White, H.D.; McCain, K.W. - Visualising a science: an author co-citation analysis of information science, 1972-1995. O In: Journal of the American Society for Information Science, 49(1998)4, p. 327-355.

12 Een voorbeeld van bibliometrie toegepast op de informatiewetenschap zelf is (naast het

hierboven geciteerde artikel van White en McCain): Tijssen, R.J.W.; Leeuwen, Th.N. - Bibliometrische beschrijving van Nederlandse informatiewetenschap in een internationaal perspectief. - Leiden: CWTS, 1995.

- 13 Het begrip 'gevraagde kennis' wordt in het algemeen geoperationaliseerd als 'die documenten die relevant zijn voor de gestelde vraag'. Bij het bepalen van de kwaliteit van een IR-systeem speelt het begrip 'relevantie' dan ook een sleutelrol. De betekenis van dit begrip is echter problematisch, zie: Mizzaro, S. - Relevance: the whole history. - In: Journal of the American Society for Information Science, 48(1977)9, p. 810-832. Ernstig is m.i. vooral het feit dat relevantie (zoals in het meeste onderzoek gedefinieerd en bepaald) weinig zegt over bruikbaarheid en betrouwbaarheid van de verkregen informatie. Zie over dit laatste o.a.: Wilson, P. - Second-hand knowledge: an enquiry into cognitive authority. - Westport: Greenwood Press, 1983, p. 171- en Fox, Ch. J. - Information and misinformation: an investigation into the notions of information, misinformation, informing and misinforming. - Westport: Greenwood Press, 1983.

- 14 De term 'information retrieval' is voor het eerst gebruikt in 1951. Zie: Mooers, C. - Zatcoding applied to mechanical organization of knowledge. - In: American documentation, 2(1951)1, p. 20-32. De eerste poging tot een integrale behandeling van het onderwerp was: Becker, J. ; Hayes, R.M. - Information storage and retrieval: tools, elements, theories. - New York: Wiley, 1963. De klassieke teksten op dit terrein zijn:

- Lancaster, F.W. - Information retrieval systems: characteristics, testing and evaluation. - 2nd. ed. - New York: Wiley, 1979. en Lancaster, F.W.; Warner, A.J. - Information retrieval today. - Rev. ed. - Arlington: Information resources Press, 1993.
  - Rijsbergen, C.J. van - Information retrieval. - 2nd ed. - London: Butterworths, 1979.
  - Salton, G. - Introduction to modern information retrieval. - New York: McGraw-Hill, 1983 en Salton, G. - Automatic text processing: the transformation, analysis and retrieval of information by computer. - Reading, Mass.: Addison-Wesley, 1989.
- Moderne teksten zijn:
- Kowalski, G. - Information retrieval systems: theory and implementation. - Boston: Kluwer Academic, 1997.
  - Schäuble, P. - Multimedia information retrieval: content-based information retrieval from large text and audio databases. - Boston: Kluwer Academic, 1997.

Een overzicht van klassieke teksten geeft: Sparck Jones, K.; Willett, P. - Readings in information retrieval. - San Francisco: Morgan Kaufmann, 1997.

Voor het speciale terrein van multimedia zie ook: Chiaramella, Y.; Mulhem, P.; Fourel, F. - A model for multimedia information retrieval. - University of Glasgow, 1996. - Technical report, FERMI ESPRIT BRA 8134.

- 15 Een andere formulering van het 'information retrieval vraagstuk' luidt: Op welke wijze kan men relevante informatie onderscheiden van niet-relevante informatie met betrekking tot een zekere informatiebehoefte? Zie: Huibers, T.W.C.; Linden, B. van; Bruza, P.D. - Een theorie voor het bestuderen van information retrieval modellen. - In: Noordman, L.G.M.; Vroomen, W.A.M. (red.) - Informatiewetenschap 1994: wetenschappelijke bijdragen aan de derde STINFON-conferentie. - Tilburg, 16 december 1994, blz. 85-102.

- 16 Deze anomalieën van de informatieketen worden natuurlijk veroorzaakt door de anomalieën van informatie zelf. Deze zijn: 'hoeveel je er ook van neemt, er gaat nooit iets af' en 'hoe meer je ervan gebruikt, des te meer komt er bij'. De hoofdanomalie die de grondslag van de informatiewetenschap vormt, luidt evenwel: 'er is te veel informatie'. Voor de inrichting van de informatieketen is het verder van belang dat informatie als eigenschap heeft dat het zeer eenvoudig vernietigd kan worden. Zie over deze zaken het afscheidscollege van een mijner voorgangers: Kouwenhoven, A.O. - De anomalie van informatie: afscheidscollege gehouden in de Aula van de Universiteit van Amsterdam op 25 maart 1992.- Amsterdam, 1992. - (BBI-reeks; nr. 1).

- 17 Voor een overzicht van de linguïstische benadering van information retrieval, zie:
- Smeaton, A.F. - Using NLP or NLP resources for information retrieval tasks. - In: Strzalkowski, T. - (ed.) - Natural language information retrieval. - Kluwer Academic Publishers (in press). - [<http://www.compapp.dcu.ie/~asmeaton/NLP-IR-Book.ps>].
  - Smeaton, A.F. - Information retrieval: still butting heads with natural language processing? - In: Pazienza, M.T. (Ed.) - Information extraction: a multidisciplinary approach to an emerging information technology, International Summer School, SCIE-97, Frascati, Italy, 14-18, 1997. Lecture notes in computer science, Vol. 1299, Springer, 1997, p. 115-138. - [<http://www.compapp.dcu.ie/~asmeaton/pubs/PAIF-Summary.ps>].
- 18 Vanuit de m.i. beperkte visie dat informatiewetenschap ongeveer hetzelfde is als 'information retrieval' kan men op goede gronden redeneren dat informatiewetenschap ongeveer hetzelfde is als kennistechnologie. Zie: Vet, P.E. van der; Mars, N.J.I. - Kennistechnologie als voortzetting van informatiewetenschap met andere middelen. - In: Kempen, G. ; Vroomen, P. de (red.) - Informatiewetenschap 1991: wetenschappelijke bijdragen aan de eerste STINFON-conferentie, Nijmegen, 18 december 1991, blz. 285-295.
- 19 De klassieke IR-modellen zijn gebaseerd op een abstract begrip van de zoekvraag, d.w.z. geabstraheerd van de gebruiker die de vraag stelt. Bij modellen gebaseerd op interactieve feedback wordt de gebruiker onderdeel van het systeem, en is het resultaat van de zoekactie niet uitsluitend afhankelijk van de zoekvraag, maar ook van de interactie tijdens het zoekproces. Om die reden kan het 'interactive feedback model' worden beschouwd als fundamenteel anders dan het klassieke IR-model. Zie bijvoorbeeld: Spink, A. - Information science: a third feedback framework. - In: Journal of the American Society for Information Science, 48(1977)8, p. 728-740.
- 20 Bellardo Hahn, T. - Text retrieval online: historical perspective on web search engines. - In: Bulletin of the American Society for Information Science, 24(1998)4, p. 7-10.
- 21 Zie de verslagen van de TREC conferenties waarin nieuwe retrievaltechnieken aan gestandaardiseerde tests worden onderworpen. [<http://trec.nist.gov/>]. Over de bijdrage van linguïstische technieken ('natural language information retrieval') wordt gerapporteerd in: Strzalkowski, T.; Lin, F.; Perez-Carballo, J. - Natural language information retrieval TREC-6 Report, [<http://trec.nist.gov/pubs/trec6/papers/ge.ps>].
- 22 Vooral onder invloed van hypertextsystemen als het WWW is er tegenwoordig meer aandacht voor het verband tussen 'retrieval' en 'browsen'. Een term die deze twee begrippen omsluit, is 'information disclosure'. Zie: Bruza, P.D. - Stratified information disclosure: a synthesis between hypermedia and information retrieval. - Proefschrift Katholieke Universiteit Nijmegen. - Amsterdam: Thesis Publishers, 1993.
- 23 Van der Vet en Mars wijzen er terecht op dat ontsluiten door middel van termen (kort gezegd de door de auteur gehanteerde taal) inferieur is aan het gebruik van een ontologie, d.w.z. een gestructureerd systeem van concepten en relaties waarin het IR-systeem de door de auteur vastgelegde kennis representeert. Zie: Vet, P.E. van der; Mars, N.J.I. - De kwaliteit van kennis. - In: Mars, N.J.I. (red.) - Informatiewetenschap 1992: wetenschappelijke bijdragen aan de tweede STINFON-conferentie, Enschede, 18 december 1992, blz. 77-88. Een vertaling vanuit de *tekst* van het document naar de ontologie blijft echter noodzakelijk, en dat is - onder meer gezien de ambiguïteit van de taal - nog altijd problematisch.
- 24 Ik gebruik het woord 'betekenis' hier in de zin van het Engelstalige woord 'aboutness', datgene waarover de informatievraag van de gebruiker - en als afgeleide daarvan de inhoud van het document - handelt. In het Nederlands wordt 'aboutness' gewoonlijk uitgedrukt in zinnen als: 'ik zoek informatie over X'. In het algemeen wordt X geformuleerd in termen van het gebruikersdomein.
- 25 Deze 'gebruikersgerichte' versus 'documentgerichte' benadering is door Jan de Vuijst treffend uitgedrukt in zijn uitgangspunt: de voorspelbaarheid van de gebruikswaarde van informatie, *gegeven de gebruiker*. Zie: Vuijst, J. de - De behoefte aan informatie: commentaar op de

informatiemaatschappij. - Tilburg: Tilburg University Press, 1994, blz. 8. Zie ook: Zinman, J. - Public knowledge: the social dimension of science. - London: Cambridge University Press, 1968: *'When a scientist consults an article in a back number of a scientific journal, he (sic) is not seeking to know what the author happened to be thinking about at that particular epoch; he is looking for evidence as to what he himself should be thinking now on that subject'* (p. 103).

26 In veel projecten op het gebied van de (linguïstische) information retrieval wordt gewerkt met kennisgeoriënteerde methoden en technieken. Hierbij wordt echter steeds gewerkt op basis van het (relatief statische) kennisdomein waarop de documentverzameling betrekking heeft. Ik pleit hier voor een benadering die gebaseerd is op (zo mogelijk dynamische) modellering van het kennisdomein en de selectiecriteria van de gebruiker. In het licht van dat kennisdomein dient het retrievalsysteem de documentverzameling te interpreteren. Zie bijvoorbeeld: Wang, P.; Soergel, D. - A cognitive model of document use during a research project, Study 1: document selection. - In: Journal of the American Society for Information Science, 49(1998)2, p. 115-133.

27 Hiervoor is onder meer onderzoek op het gebied van 'information filtering' van belang, waarbij de betekenis van in een systeem binnenkomende informatieobjecten dynamisch, in 'real time' wordt bepaald. Zie onder meer:

- Belkin, N.J.; Croft, W.B. - Information filtering and information retrieval: two sides of the same coin? - In: Communications of the ACM, 35(1992)12, p. 29-38.
- Faloutsos, Chr.; Oard, D. - A survey of information retrieval and filtering methods. - Unpublished paper, University of Maryland. - [<http://www.enee.umd.edu/medlab/filter/papers/survey.ps>].
- Shapira, B.; Hanani, U.; Raveh, A.; Shoval, P. - Information filtering: a new two-phase model using stereotypic user profiling. - In: Journal of intelligent information systems, 8 (1997), p. 155-165.
- Sorensen, H.; Smeaton, A. - An overview of information filtering. - Dublin City University, Working Paper Series, CA-2097.  
Van wezenlijk belang zijn ook het modelleren van de gebruiker, kennis van informatiezoekgedrag en technieken voor gebruikersfeedback - in feite het hele domein van de 'cognitive approach'. Zie onder meer:
- Chen, H. - Machine learning for information retrieval: neural networks, symbolic learning and genetic algorithms. - Unpublished paper, Tucson, University of Arizona, 1994. - [<http://ai.bpa.arizona.edu/papers/PS/mlir93.ps>].
- Stadnyk, I.; Kass, R. - Modeling users' interests in information filters. - In: Communications of the ACM, 35(1992)12, p. 49-50.
- Spink, A. & Losee, R. M. - Feedback in information retrieval. - In: Annual Review of Information Science and Technology, 31(1996), 33-78.
- Wilson, T.; Walsh, Ch. - Information behaviour: an interdisciplinary perspective. - Boston Spa: British Library Research and Innovation Centre, 1996. - (British Library Research and Innovation Report ; 10). - [<http://www.shef.ac.uk/~is/publications/prelim.html>]. Deze publicatie bevat een uitgebreide bibliografie. Zie ook het onderzoeksproject van Wilson, Ellis en Ford: 'Uncertainty in information seeking' [<http://www.shef.ac.uk/~is/research/imrg/uncerty.html>].

28 Belkin, N.J. - The cognitive viewpoint in information science. - In: Journal of information science, 16(1990), p. 11-15. Zie ook: Ingwersen, P. - Cognitive perspectives of information retrieval interaction. - In: Journal of documentation, 52(1996)1, p. 3-50.

29 Brookes, B.C. - The foundations of information science, part I: philosophical aspects. - In: Journal of information science, 2(1980), p. 125-133.

30 Zie: Vickery, B.C. - Knowledge structures in information retrieval. - In: Journal of documentation, 48(1992)3, p. 326-328.

31 Ingwersen, P. - The human approach to information science and management. - In: Journal of



information science, 29(1994)3, p. 197-208.

Uit een advertentietekst voor een hoogleraar informatiewetenschap aan de Faculteit der Sociale Wetenschappen van de Universiteit van Tampere, Finland, 1992.

Information Engineering (IE) is een concept dat is geïntroduceerd door de EU in de context van het Telematics Applications Programme; zie voor dit programma [<http://www.echo.lu/ie/en/iehome.html>] en voor het IE-curriculum [<http://www.echo.lu/ie/en/euroiemaster.html>].

De vraag wat we precies onder ‘kennis’ moeten verstaan in de context van organisaties is buitengewoon interessant, en beslist niet eenduidig te beantwoorden. Zie bijvoorbeeld: Essers, J.; Schreinemakers, J. - Nonaka's subjectivist conception of knowledge in corporate knowledge management. - In: Knowledge organization, 24(1997)1, p. 24-32.

Die zich sinds enige tijd ten onrechte graag ‘informatieprofessional’ noemen. Ten onrechte, omdat daarmee wordt gesuggereerd dat er naast documentaire informatie geen andere informatie zou zijn waarmee men zich professioneel kan bezighouden. Het tegendeel is natuurlijk waar.

Voor een andere visie op kennismanagement zie: Stoffle, C.J. - The emergence of education and knowledge management as major functions of the digital library. - Follett lecture, University of Wales Cardiff, Wales, 13th November, 1996. - [<http://ukoln.bath.ac.uk/follett/stoffle/paper.html>].

Het zelf uitgeven van eigen werk is allesbehalve nieuw, en zeker niet voorbehouden aan tweederangs auteurs. Zo is bijvoorbeeld van Terentius (195-159 v. Chr.) bekend dat hij zijn eigen werk vermenigvuldigde en verspreidde.

Onder ‘digitaal tijdschrift’ verstaan we een zelfstandige digitale publicatie die niet de gedigitaliseerde versie is van een bestaand gedrukt tijdschrift. Een bekend voorbeeld is: Psycology - [<http://www.cogsci.soton.ac.uk/psycology/>]. De digitale versies van gedrukte tijdschriften, zoals die door Elsevier, Wolters Kluwer en andere uitgevers worden uitgebracht, rekenen we dus niet tot de digitale tijdschriften, al zijn het wel digitaal gedistribueerde tijdschriften.

De discussie over de rol van wetenschappelijke uitgevers en de mogelijkheden voor ‘self-publishing’ heeft een sterke impuls gekregen door de publicatie ‘Scholarly journals at the crossroads: a subversive proposal for electronic publishing’ onder redactie van Ann Okerson & James O'Donnell. (Washington, DC, Association of Research Libraries, June 1995). Met name de initiator van deze (geheel uit e-mailcorrespondentie bestaande) publicatie, Steven Harnad, heeft dit onderwerp in tal van publicaties verder uitgedragen. Zijn stelling is (1) dat auteurs het recht moeten hebben om hun manuscripten gratis en vóór publicatie via het Internet te verspreiden, (2) dat de kosten van het publiceren door de auteur (via ‘page charging’) moeten worden betaald en (3) dat digitale netwerken dat zonder meer mogelijk maken. Voor een overzicht zie: [<http://cogsci.soton.ac.uk/~harnad/intpub.html>]. Zijn meest recente bijdrage (‘The paper house of cards’) verscheen in het elektronische tijdschrift Ariadne (issue8, maart 1997). Zie: [<http://www.ariadne.ac.uk/issue8/harnad/>].

De bekendste voorbeelden van pre-print archieven zijn:

- CogPrints Cognitive Science Eprint Archive. - [<http://cogprints.soton.ac.uk/>]
- Los Alamos National Laboratory Physics Service. - [<http://xxx.lanl.gov/>].

Voor een bespreking van randvoorwaarden voor succesvolle invoering van pre-print servers zie: Requirements on first publication archives. - In: ETAI Electronic transactions on artificial intelligence, 1997. - [<http://www.ida.liu.se/~erisa/org/spai/ntd97/04/online.html>].

Over het verschijnsel ‘digitale publicaties’ bestaat inmiddels een zeer uitgebreide literatuur. Een van de eersten die de vraag aan de orde heeft gesteld of digitaal publiceren tot het verdwijnen van het traditionele wetenschappelijke tijdschrift zal leiden, is Andrew Odlyzko met zijn beroemde artikel ‘Tragic loss or good riddance: the impending demise of traditional scholarly journals’ (in: International journal of human-computer studies, 42(1995), pp. 71-122,

en in het elektronische tijdschrift J. Univ. Comp. Sci., pilot issue, 1994. Zie:  
[<http://www.research.att.com/~amo/doc/eworld.html>]. In een recente bijdrage wijst hij erop dat de wetenschappelijke communicatie zich slechts heel langzaam aanpast aan de technologische mogelijkheden. Zie: Odlyzko, A.M. - The slow evolution of electronic publishing. - In: Meadows, A.J.; Rowlands, F. (eds). - Electronic publishing: new models and opportunities. - ICC Press, 1997. - [<http://www.research.att.com/~amo/doc/slow.evolution.txt>].

Voorbeelden zijn Science Direct van Elsevier [<http://www.sciencedirect.com/>] en LINK van Springer Verlag [<http://link.springer.de/>].

Mackenzie Owen, J.S. - Licenties en digitale archivering. - In: Informatie professional, 1(1997) 2, p. 39-41.

Ik heb het hier over 'traditionele' bibliotheken die hun dienstverlening in toenemende mate baseren op digitale bronnen. In de literatuur wordt ook in een andere betekenis over 'digital libraries' gesproken, namelijk in de zin van verzamelingen digitale informatieobjecten die via een netwerk met elkaar interacteren. Deze vorm van 'digital libraries' is nauw gelieerd aan ontwikkelingen op het gebied van elektronische zakelijke transacties ('electronic commerce') via het Internet, en heeft in principe weinig van doen met ontwikkelingen binnen het geïnstitutionaliseerde bibliotheekwezen. Zie: Adam, N.; Yesha Y. et al. - Strategic directions in electronic commerce and digital libraries: towards a digital agora.. - In: ACM computing surveys, 28(1996)4, p. 818-835. Overigens vertonen Amerikaanse bibliotheken (o.a. via het 'Digital Libraries' project), in tegenstelling tot het bibliotheekwezen in Europa, een duidelijke tendens om zich te ontwikkelen in de hier genoemde richting. Zie:

[[http://www.cise.nsf.gov/iis/dli\\_home.html](http://www.cise.nsf.gov/iis/dli_home.html)]

Mackenzie Owen, J.S. - Een toekomst voor de bibliotheek: op weg naar de netwerkbibliotheek. - In: Informatie professional, 1(1997)6, blz. 14-17.

Zinman, J. op.cit., p. 102.

Zoals bekend is er groeiende kritiek op de rol van uitgevers, die soms meer gericht lijken te zijn op winstmaximalisatie dan op bevordering van de wetenschappelijke communicatie. Deze kritiek is niet alleen recent. Zie bijvoorbeeld: Horowitz, I.L. - Communicating ideas: the crisis of publishing in a post-industrial society. - New York: Oxford University Press, 1986.

Het begrip 'digitale archivering' gebruik ik hier, in navolging van vele anderen, om het voor de langetermijn bewaren en toegankelijk houden van digitale *publicaties* aan te duiden. In de archiefwereld wordt uiteraard ook gesproken over digitale archivering, en in beide domeinen vinden we gelijksoortige technische vraagstukken. Zie bijvoorbeeld: Dollar, Ch.M. - Ensuring access over time to authentic electronic records: strategy, alternatives and best practices. - University of British Columbia, November 1997 (draft paper). Zie ook de verwijzing naar Digitale Duurzaamheid hierna.

In Nederland is de Koninklijke Bibliotheek actief op het gebied van de digitale archivering van in Nederland verschenen digitale publicaties via het Depot van Nederlandse Elektronische Publicaties (DNEP). (Zie: Noordermeer, T. - Depot van Nederlandse Elektronische Publicaties. In: Informatie professional, 2(1998)2, p. 22-24). Voor een overzicht zie: Hoare, P. - Legal deposit of electronic publications and other non-print material: an international overview. - In: Alexandria, 9(1997)1, p. 59-79. Verder is in Nederland een Programmabureau Digitale Duurzaamheid opgezet dat tot doel heeft een duurzame toegankelijkheid en beschikbaarheid van digitale overheidsdocumenten te waarborgen en zo te verzekeren dat de overheid ook in het digitale tijdperk betrouwbaar kan functioneren. Zie: [<http://www.archief.nl/digiduur/>].

Andere organisaties op het gebied van digitale archivering zijn:

- European Commission on Preservation and Access - European Preservation Information Center. - [<http://www.knaw.nl/ecpa/ecpatex/home.htm>].
- Commission on Preservation and Access (USA). - [<http://www.clir.org/cpa/>].

Dat blijkt in Nederland waar er, in tegenstelling tot de meeste andere Europese landen, geen wettelijke depotregeling is die uitgevers verplicht om een of meer exemplaren van hun

publicaties aan de nationale bibliotheek af te staan. In Nederland gebeurt dat op vrijwillige basis, en volgens mededeling van de Koninklijke Bibliotheek leidt dat tot een hogere dekking van de nationale productie dan in veel landen met een wettelijke regeling.

50 Zie: Auteursrechtelijke aspecten van preserving van elektronische publicaties. - Amsterdam: Instituut voor Informatierecht, februari 1998.

51 De term 'verdoeken' is afkomstig van Dr. W. van Drimmelen, Bibliothecaris van de Koninklijke Bibliotheek. Verdoeken geeft precies aan waar het om gaat: conserveren van het origineel, zonder dat origineel aan te tasten.

52 De toenemende macht van de consument strekt zich ook uit tot de *structuur* van de tekst: nieuwe media zijn niet meer gebaseerd op een lineaire ordening, maar maken het de lezer gemakkelijk om een eigen leespad door de tekst te creëren, gebruik makend van hyperlinks en andere 'navigatiemiddelen'.

53 Mackenzie Owen, J.S. - Kennis, betoog en conversatie: veranderingen in het gebruik van de informatieketen. - In: Marks, A.F. (red.) - Sociaal-wetenschappelijke informatie en kennisvorming in onderzoek en beroep. - Amsterdam: SWIDOC, 1989, blz. 39-47 (gevolgd door een commentaar van J.M. van Oorschot en een naschrift van de auteur, blz. 49-54).

54 Mackenzie Owen, J.S. - Wetenschapper en informatiemaatschappij. - In: I&I Informatie en informatiebeleid, 5(1987)2, blz. 70-78. (NB: Het schema op blz. 71 dient te worden vervangen door dat op blz. 65).